PSY652, Unit 2, In class activity

In this activity you will build on the Fandango R NOTEBOOK from Unit 1.

Dataframe: In Dropbox Unit1: fandango.csv

Description of Variables: In Dropbox Unit1: Fandango Dataframe Variable Descriptions

1. Open the Fandango Notebook. At the bottom of the notebook (after your simple graph), create a first level header and call it: A collection of graphs (be sure to put a single hashtag before this phrase).
2. Now create a second level header called: A density plot of the IMDb ratings (be sure to put two hashtags before this phrase). And then add a code chunk just below this header (INSERT > R). Create a density plot of the IMDb ratings (the variable name is imdb). Look up your favorite color in the Rcolors pdf (Dropbox\PSY\_RMstudent\_materials\Other Resources), and fill the density plot with this color.
   1. Once created, rather than using fill = , use color = . What happens?
   2. Enhance this density plot, give the x-axis the title of “IMDb user ratings” and then choose an informative title for your graph.
   3. Change the graph title to something really long – I chose "A density plot of the IMDb user ratings for movies, compiled by fivethirtyeight in mid 2015" – execute the graph and notice that the title runs off the graph. You can make a return by adding \n to the spot where you want a return – so for example if I want a return after the comma but before “compiled”, I would have: "A density plot of the IMDb user ratings for movies, **\n**compiled by fivethirtyeight in mid 2015". Apply this fix to your graph.
3. Add a new code chunk – give this code chunk a second level header called: A density plot of the IMDb ratings by film year. Copy your final density plot from the previous step, remove the fill assignment in the geom\_density line (leave the ellipses), and now add: group = year, to the aes argument.
   1. Modify this graph, after group = year, add fill = year.
   2. Notice the legend. Because year is a numeric variable, R treats it as such – but it’s discrete. To remedy this situation change fill = year to fill = factor(year). Notice the change in the legend.
4. Add a new code chunk – give this code chunk a second level header called: A bar chart of Fandango star ratings. Create a bar chart to display the number of films that received each star level (the variable is called fandango\_stars). Give your chart an informative title.
5. Add a new code chunk – give this code chunk a second level header called: A boxplot of IMDb ratings by fandango stars. Plot the imbd score on the y-axis, and fandango\_stars on the x-axis. Note that ggplot requires that the x variable for a box plot is categorical, so map x as follows: x = factor(fandango\_stars). Give your chart an informative title.
6. Add a new code chunk – give this code chunk a second level header called: A scatterplot of the relationship number of IMDb votes and the IMDb rating. Make a scatterplot to present the relationship between number of IMDb votes (imdb\_user\_vote\_count) and IMDb score (imdb). Add a best fine line using the geom\_smooth argument and the lm option for linear model. Give your graph an informative title, and give both the x and y axis informative labels.
7. Once everything is created, Restart R and Run All Chunks, and then preview you notebook. Save your notebook and exit RStudio.



